The Effect of Greenspace on Physical Activity and Social Cohesion in Baltimore, MD

Elizabeth P. Tong[*†]

April 20, 2020

Abstract

Given increasing trends in urbanization, there is an increasing need for sustainable city design solutions that will improve citizens' health, well-being, and efficiency in high-density environments. One large contributor to urban well-being is greenspace. Thus, this paper seeks to analyze the effect of greenspace on physical activity rates and social cohesion in Baltimore, MD. Using Landsat 4-5 imagery, NDVI calculations were used to measure greenspace as the percentage of healthy vegetation within zip codes and census tracts. Repeated cross-sectional data at each geographic area is available in the Baltimore Ecosystem Study Telephone Survey from 1999 to 2011. As noted in Baltimore's Sustainability Plan, the city of Baltimore intentionally targets increased greenspace for low-income, African-American, and minority residents, which means greenspace is non-random. The fixed effects model accounts for this endogeneity and is applied at each geographic level to control for unobserved neighborhood characteristics. The results find a statistically significant positive relationship between greenspace and social cohesion, but inconclusive results on the relationship between greenspace and physical activity.

Keywords: greenspace, physical activity, exercise, neighborhood, social cohesion, urban space, health, Baltimore

JEL Codes: R58, I40, H4

[*]University of North Carolina at Chapel Hill, Department of Economics, David Guilkey
[†]University of North Carolina at Chapel Hill, Department of Economics, Jane Fruehwirth

1 Introduction

The United Nations predicts that by 2050, 68% of the global population will reside in cities, shifting 2.5 billion people from rural to urban areas. In order to prepare cities to successfully manage this urban growth in the face of diminishing natural resources, increasing population density, and smaller geographic constraints, a deeper understanding of best practice sustainable city design is necessary. While other studies are addressing the challenges of affordable housing, efficient transportation networks, carbon neutral sources of energy, and human-scale infrastructure, more research is needed to inform policy decisions that influence the health and well-being of residents.

In dense urban areas, proper management of public space, specifically urban green space – parks, forests, streams, rivers, and gardens – can influence the built environment and have a significant impact on the physical activity rates, but an increasing number of studies note its benefits to ecology, social cohesion, and well-being. Previous studies also show that most urban green space is distributed in wealthy neighborhoods (Heynen, Perkins & Roy, 2006). This suggests that a barrier exists that prevents low-income, African-American, or minority communities from reaping benefits associated with urban greening. Additionally, research on the relationship between greenspace on physical and mental health find conflicting results (Lachowycz & Jones, 2010). Thus, I present new estimates on this research question, the effect of greenspace on physical activity rates and social cohesion, focused in Baltimore, Maryland.

My results build off traditional theories on the impacts of neighborhood characteristics, assessing how (1) accessibility of amenities (2) size or quality of amenities, impact physical activity and social cohesion. This paper uses a framework from health economics literature that incorporates characteristics of the built environment into an individual's demand for health model. Data for the empirical analysis will use satellite data from Landsat 4-5 to generate Normalized Difference Vegetation Index (NDVI) calculations of healthy vegetation in ArcGIS and associate it with data on outdoor physical activity from the Baltimore

Ecosystem Study Telephone Survey (BESTS).

This dataset is combined with data from American Community Survey, Baltimore Open City data, and Baltimore Ecosystem Study geographic files to represent the built environment, ultimately building to 1425 households across 1999-2011. By applying a fixed-effects model, at the zip code and census tract levels, I provide estimates on a statistically significant causal relationship between greenspace and social cohesion and inconclusive results between greenspace and physical activity rates.

I introduce my methodology in the following sections of the paper. In Section 2, I review previous literature explored across public health, urban planning, and economics journals. Section 3 introduces a theoretical model that applies neighborhood characteristics into the demand for health. Section 4 explores the data I use, and Section 5 explains the empirical model. Section 6 and 7 present the results and discussion of its implications. Finally, Section 8 concludes the paper with larger policy implications.

2 Literature Review

A vast body of literature exists on the relationship between greenspace and individual health outcomes. Both greenspace and health are incredibly broad topics; thus, many papers identify very specific definitions of either variable. Greenspace can be broken down into two categories, availability of greenspace compared to quality of greenspace. Similarly, health is frequently broken down into two categories, physical and mental health, with physical health often being represented by physical activity rates and mental health often associated measured by stress and community social cohesion levels (de Vries et al., 2013). This paper contributes to the literature by estimating a causal relationship between greenspace and physical and mental health, more rigorously in sample size and methodology than previous studies.

2.1 Greenspace and Physical Health

The literature on greenspace and physical health is divided into two strands: the first focuses on the characteristics of the built environment that effect physical activity and the second analyzes socioeconomic variation in physical activity. A comprehensive review of 100 studies by Barnett et al. (2017) finds the following characteristics – walkability, safety from crime, access to destination and services, recreational facilities, parks and open spaces, commercial destinations, and greenery – to all be statistically significant components of physical activity, listed in order by strength of evidence. These characteristics are taken into account in the analysis of this paper and outside sources of data that represent the built environment are utilized where available.

Most of the literature on the effects of the built environment exist in public health, urban planning, and epidemiology journals. Fewer studies have looked at the effects of greenspace specifically; however, a recent synthesis of research on these topics have found that poor health effects are associated with lower availability of environmental resources (Schüle et al., 2019). One study assessing the accessibility of recreational facilities finds that households of lower socioeconomic status had reduced access to facilities, which then reduced physical activity (Gordon-Larsen et al., 2006). Their paper considers a 5-mile radius of recreational areas within a household, a boundary captures relevant physical activity facilities based on empirical evidence by Sallis et al. (1990). In my paper, I will attempt to mimic accessibility by using data with geographic boundaries at the zip code and block group level.

Further, very little literature applies econometric methods to determine the causal links between green space and physical activity and social cohesion. One analysis on vegetation and crime in Philadelphia uses spatial econometric models to conclude that the crime rate in one neighborhood is associated with crime rates in adjacent neighborhoods (Wolfe & Mennis, 2012); however, this study still does not seek to understand how vegetation is related to crime. Ghimire et al. (2017) models the effect of greenspace on adult obesity through BMI, controlling for socioeconomic and demographic differences at the individual and county level as well as state-specific heterogeneity but does not study the effect of greenspace on physical activity rates. Meyer et al. (2016) is one of the few papers in epidemiology and public health journals that have applied econometric methods, IV simultaneous equations modeling, to determine the effect of physical activity on BMI using neighborhood, in addition to traditional analysis for longitudinal data such as ordinary least squares, random-effects regression, and fixed effects regression. This is method is better than the standard IV method that accounts for unmeasured confounding, measurement error, and use of semiparametric and nonlinear estimation. This paper is answering a slightly different question, the effect of green space on physical activity rates, but applying similar econometric approaches that can determine a causal relationship.

The second strand of literature focuses on socioeconomic variation and studies its impact on physical activity. Analyzing socioeconomic characteristics is important to my research question because variables like race, income, and gender can be confounding variables that impact physical activity rates or levels of social cohesion. Saffer et al. (2011) examines the racial, ethnic, and gender differences in physical activity. Their measurement of physical activity, based on the specificity provided by the American Time Use Survey (ATUS), is able to label all types of physical activity, including work and non-work related exercise. While their specificity of physical activity is not available in the data provided in my analysis, their strategy of measuring physical activity by utilizing Metabolic Equivalents of Task (METs) is applied in my paper. Further, Lachowycz & Jones (2013) present a new theoretical framework assessing greenspace and health, suggesting that the relationship's effect on varying demographic characteristics is important for understanding how mediators of greenspace and health – the use of greenspace and perceptions of living environment – operate. The results and discussion in my analysis will build upon this second strand of research in an attempt to better understand the overall relationship between greenspace and health and its variations in different demographic breakdowns.

2.2 Greenspace and Mental Health

A much smaller body of literature exists on the association between greenspace and mental health. A systematic review on greenspace and mental health finds no consensus or statistical significance (Lachowycz & Jones, 2011). Some papers have even found a negligible or negative relationship (Astell-Burt et al., 2013). Given the difficulty of quantifying well-being through survey responses and the broad definition of mental health, variations in conclusion are common. Finer measurements of mental health, like social cohesion, which is defined as an individual's sense of belonging, trust, and acceptance in a community, are also important measures of mental health. The World Health Organization notes poor mental health as one of the largest challenges facing the European region. Additionally, they find social cohesion, or good relationships within the home, school, and neighborhood, as essential to promoting the health and well-being in children. Marmot & Wilkinson (1999) notes that social support improves health through two key mechanisms. First, the positive, direct effects of social support on health, from a lack of social isolation; second, the moderating effect social support has on buffering acute and chronic stressors on health. These hypotheses are backed by public health studies that find strong relationships between neighborhood social cohesion and more traditional measures of mental health, like depression (Echeverria et al., 2008). Because social cohesion is a mediator of mental health and a measurement of the built environment, a subset of studies that analyze the relationship between social cohesion and greenspace find a positive relationship (Beyer et al., 2014, Nutsford et al., 2013, Maas et al., 2009).

A study by Hong et al. (2018) examining the association between greenspace and social capital on older adults, mediated through an individual's perception of safety, is used as the focal paper and point of comparison for my analysis of greenspace on social cohesion. The results of Hong find a statistically significant relationship between greenspace on social cohesion, using census block groups as a random effects cluster variable. My paper assesses the same research question in a more robust methodology. The study by Hong is a cross-sectional sample distributed throughout Seattle and Baltimore with a total of 647 participants and defines greenspace in three ways – the presence of natural sights, presence of street trees, and park accessibility. The presence of street trees and natural sights were derived from survey responses from the Neighborhood Environment Walkability Scale (NEWS) rating on a four-point scale, from strongly agree to disagree, asking respondents how strongly they agreed that there are many trees in their neighborhoods and there are many attractive natural sights in their neighborhood. Thus, I contribute to the literature assessing the relationship between greenspace and social cohesion by providing a more quantitative and objective measure of greenspace with a larger sample size, applying causal analysis. This is crucial, since most studies assessing social cohesion still struggle to establish a causal relationship (Lee & Maheswaran, 2011).

2.3 Contribution to the Literature

The studies highlighted in this literature review do not determine causality because of residential self-selection, unobserved heterogeneity, and reverse causality. Many studies have applied a variety of methods to address non-random neighborhood mobility and its effects on estimation bias on health – introducing larger set of covariates, difference in difference, panel models, and instrumental variables (Do & Finch, 2008, Eid et al., 2008). Past studies analyzing residential self-selection have found that financial resources and household structures influence where people are financially capable, or decide to live (Clark & Ledwith, 2007) and can thus have an impact on outcome variables tied to neighborhood environment. A study uses economic analysis to find the determinants of obesity takes neighborhood selection into account – certain characteristics and resources of individuals that choose to live in a given area – by using choice variable for residential location decision (Morales et al., 2016). A paper on neighborhood SES predictors of physical activity highlight that these choice variables need to account for characteristics like marriage and children, in addition to individual SES into account when controlling for residential selection (Boone-Heinonen et al., 2011).

Taking these above methodologies into account, this paper contributes to the literature by providing a robust assessment of the causal relationship between greenspace availability and health. Using an objective measure of greenspace, my analysis applies fixed effects model to address residential selection in the context of greenspace and its effects on physical activity and social cohesion at the zip code and block group level.

3 Theoretical Model

The theoretical model for this paper is based on the demand for health model first created by Grossman (1972). This model is further modified to account for non-medical commodities, like environmental factors and greenspace. This section provides a overview of how my research question - the effect of greenspace on physical activity and social cohesion - falls into the field of health economics by measuring demand for health. Although the goal of the research is to calculate the improvement on overall individual health, the dataset in this paper does not include a measure of overall health. Instead, the results of this paper show how greenspace affects the mediators of health, physical activity and social cohesion. Regardless of the availability of the measurement for overall health in the dataset used in this analysis, the model below provides a framework that maximizes health while taking neighborhood environmental factors into account. Future studies that use datasets which include baseline measurements of overall health, would be able to be able to apply the following adaptation of the traditional health economics model to calculate impact on overall health while taking into account environmental factors.

This demand for health model is based off Grossman's seminal model. Grossman's model differentiates itself from previous health models by allowing individuals to weigh their own costs and benefits to health in order to calculate and offset their own loss. The model provides two major contributions, the incorporation of (1) an intertemporal utility function where utility depends on the flow of healthy days (h_t), and (2) the stock of health available in a given time period (H_t). The total health stock is also composed of the consumption of other goods, which are defined by price and time. Changes in health stock are then composed of changes in the gross investments of health and the depreciation of health stock, caused by age.

The Grossman model fails to include non-medical commodities as health investments that compete with medical services for an individual's monetary resources and time. These non-medical commodities like smoking tobacco, can provide utility in time t, but can decrease the number of healthy days in time $t+1$. Similarly, physical activity, may provide disutility in time t, but can provide long-term benefits to health, increasing the number of healthy days in time $t+1$. Both these examples, however, can also be addictive. Thus, utility retrieved in time t may be higher for an addict to exercise, thus outweighing the cost and loss to health. All these non-medical commodities compete with medical care in order to impact health stock. Thus, non-medical commodities specific to this paper, like neighborhood characteristics, are included in the following Grossman model:

$$\Delta H = H(t_m, t_c, t_w, M_t, C_t, N; H_{t-1}, E) \tag{1}$$

where net investments in health are composed of the cost of medical service (M_t), cost of non-medical commodities (C_t), cost of environmental inputs (N), and time spent on medical services, non-medical commodities, and work. Education, (E), represents a factor for efficiency. The theory finds that individuals with greater educational attainment are better able to manage their time optimally to receive the best health output. Thus, education represents a proxy for efficiency. The above model, first compiled after reviewing thirty years of health economics literature, is used as a basis for understanding the impact of environment on demand for health (Leibowitz, 2004). In that model, Leibowitz considers environmental inputs as exogenous, beyond an individual's control; thus, the same assumptions will be made in this paper's model. Environmental characteristics not eliminated by the fixed effects analysis will be considered exogenous and represented by N. The following equation, based off Saffer et al. (2011), represents the demand for physical activity that can be attained given the data used in this paper. The analysis seeks to maximize an individuals' utility from physical activity and their demand for physical activity:

$$PA = PA(X, N; E) \tag{2}$$

where X represents socioeconomic indicators like race, sex, and income, and N still represents exogenous environmental characteristics. If the dataset used in my analysis had provided overall health, the maximized demand for physical activity would then be substituted to the following equation, the final adapted demand for health:

$$\Delta H = H(t_{pa} * MET, t_m, t_c, t_w, PA_t, M_t, C_t, N; H_{t-1}, E) \tag{3}$$

Physical activity is explicitly incorporated into the demand for health model based on the total amount spent on physical activity (PA_t) and the time spent on exercise (t_{pa}) multiplied by the Metabolic Equivalent of Task (MET). Social cohesion, which can be viewed as an aspect of mental health, is incorporated into equation 3 by influencing cost and time spent on mental health services. Effects of social cohesion can also be considered in the cost of environmental inputs based on the socioeconomic status of a neighborhood.

Equation 3 is the final adapted model that can be used to capture the effects of physical activity and social cohesion within a traditional health economics framework. Again, the Baltimore Ecosystem Study Telephone Survey (BESTS) does not include an outcome variable that measure overall health so the demand for health cannot be estimated. Instead, the results of the empirical analysis are used to determine a causal relationship between greenspace and the explanatory variables, physical activity and social cohesion.

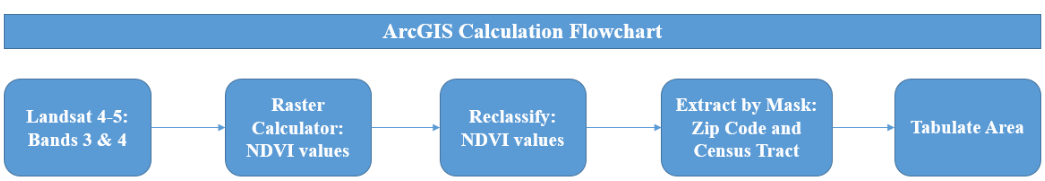
4 Data

4.1 Greenspace

The data that captures the key independent variable, greenspace, comes from US Geological Survey remote sensing satellite data. Landsat 4-5 satellite takes images from months of high vegetation coverage, May to August, over the time period 1999 to 2011, into order to estimate the amount of greenery in a geographic region. The indicator of greenery is the normalized difference vegetation index (NDVI), which is commonly used in agricultural and environmental sciences in order to evaluate plant and crop health. This value is calculated by assessing the amount of light and reflectance from Near-Infrared (NIR) light and red visible light, because healthy vegetation reflects higher amounts of NIR light. The following formula calculates NDVI values:

$$NDVI = \frac{band4 - band3}{band4 + band3} \tag{4}$$

which range from -1 to 1, with the negative values representing non-plant surfaces, values closer to zero representing unhealthy vegetation, and values closer to 1 representing healthy vegetation. Landsat 4-5 uses band 3 capture red visible light and band 4 to capture NIR light. Thus, the NDVI values determine healthy vegetation to be NDVI values over .34, or moderately to very healthy vegetation. ArcGIS Pro is the program that is used to calculate the greenspace values for each geographic area. The follow flowchart illustrates the process:



After uploading the bands, the raster calculator is used to determine the NDVI values for the geographic area. These NDVI values, are reclassified into categorical variables, from 1 to 6, as illustrated in Table 1.

NDVI Categorization (Table 1)					
Dead Plant or Inanimate Object			Unhealthy	Moderately Healthy	Very Healthy
1	2	3	4	5	6
-1 to -.67	-.66 to -.34	-.33 to 0	0 to .33	.34 to .66	.67 to 1

Healthy Vegetation – Zip code (Table 2)				
	Mean	Std Dev	Min	Max
Overall	.4581	.2461	.0685	.9761
Between		.2460	.1223	.9278
within		.0941	.1011	.6080

Extract by Mask is then used to determine NDVI values for geographic boundaries of Baltimore. Tabulate area is then used to determine the total square meters of each geographic area is then categorized into these NDVI values that represented inanimate objects, unhealthy, to very healthy plants. The average amount of greenspace value was calculated for each geographic area – Baltimore zip code and census tract level – by summing the total number of square meters of healthy vegetation over total square meters of a given area for each year:

$$Greenspace = \frac{5+6}{1+2+3+4+5+6} \tag{5}$$

represented by the greenspace variable, consists of the percentage of area in a geographic area with healthy vegetation. Tables 2 and 3 analyze how much greenspace or vegetation varies within and across both zip codes and census tracts. Table 2 indicates that the average amount of greenspace in a given zip code is 45%, however, within each zip code across time, the amount of greenspace varied by as much as 9.4%. Similarly, across zip codes across time, the amount of greenspace varied by as much as 24.6%. Table 3 indicates the amount of greenspace variation between and across census tracts.

The table above notes that smaller geographic areas tend to have a higher percentage of the area that are considered to be greenspace or healthy vegetation; however, the census

Figure 1: NDVI 1999 Baltimore

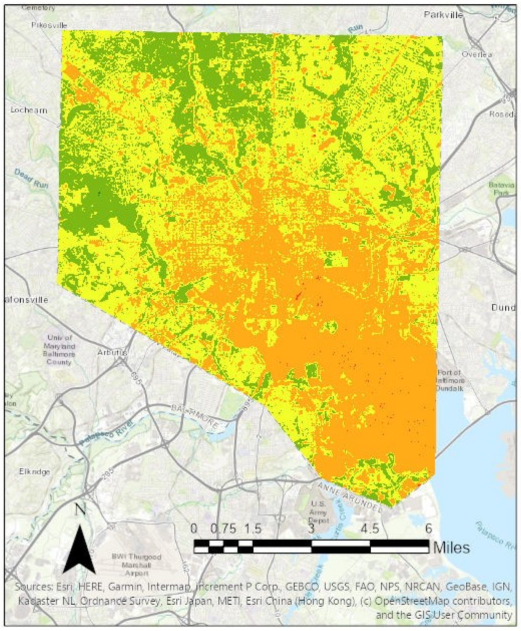
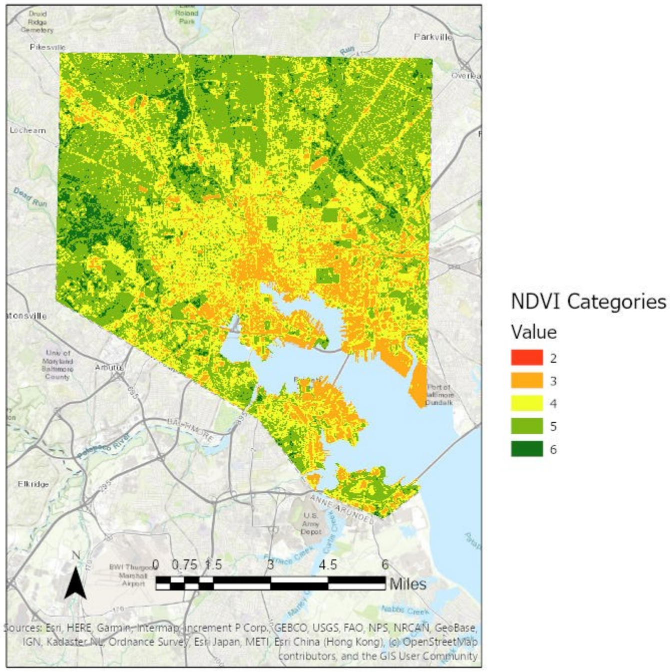


Figure 2: NDVI 2011 Baltimore



Healthy Vegetation – Census Tract (Table 3)				
	Mean	Std Dev	Min	Max
Overall	.6872	.2870	0	1
Between		.2019	.0456	1
Within		.2117	.1515	1

tract areas also have wider standard deviations, suggesting that the range of greenspace in these areas vary more so than based on zip codes. Intuitively, larger area or sample size will tend toward closer to the average amount of greenspace in Baltimore. This average amount of greenspace is comparable to other city-level statistics available on greenspace, like tree canopy cover. Organizations like Tree Baltimore cite that Baltimore's tree canopy cover is approximately 30%. Given that this statistic does not account for any vegetation other than tree canopy, the data on greenspace above is comparable. Figures 1 and 2 are a visual representation of NDVI values from 1999 and 2011, respectively.

The main dataset that is used to capture the dependent variable, physical activity rates, is the Baltimore Ecosystem Study Telephone Survey (BESTS). The Baltimore Ecosystem Study (BES) began in 1998 and was gathered by the Cary Institute of Ecosystem Studies and the University of Maryland. Participants were reached out by telephone in a random fashion. This survey is one of many datasets collected as part of this initiative. Participants were from counties across Maryland from 1999 to 2011 with a total of 17,960 observations. This data was limited to Baltimore county and Baltimore city, narrowing the sample size down to 3,229. Each household was only sampled once during the entire survey; thus this is repeated time-series cross-sectional data.

Since the dataset that contains necessary information on physical activity and social cohesion is not panel data, multilevel fixed effects cannot be used. Only fixed effects at the geographic level, which controls for unobserved characteristics of each zip code or census tract will remain constant. Thus, this analysis holds everything in a given geographic area constant except for greenspace in order to determine the area's physical activity rates and social cohesion change for each group of individuals within a geographic area.

Summary Statistics – Physical Activity (Table 4)				
Variables	Mean	Std. Dev.	Min	Max
walkFreq	144.76	133.26	0	365
bikeFreq	43.49	96.36	0	1095
swimFreq	40.92	115.15	0	2555
sportFrq	43.85	83.75	0	720
METphyaweek	9.02	9.94	0	26.67

Metabolic Equivalent of Task Values for Activities in ATUS (Table 5)	
Physical Activity	MET
Walking	3.8
Biking	8.0
Sports	5.1
Water Sports	5.22

4.2 Physical Activity

Data found on physical activity variables in BESTS mirror standard physical activity questionnaires that are applied in large-scale epidemiological surveys. The main dependent variable, physical activity, is based off the participation in activities in moderate and vigorous levels of physical activity, like walking, biking, general sporting, swimming, and boating. Questions were asked in the following format "Now, thinking about only yourself, about how often per week, month, or year do you typically [NAME ACTIVITY]" with all frequency and responses adjusted to an annual basis. Below are summary statistics for the individual activities and the corresponding cumulative physical activity variable METphyaweek.

Previous literature that studies changes in physical activity rates use a consistent form of measurement called Metabolic Equivalence of Task (MET). METs are the energy expenditure of a given activity. For example, one MET is the energy expenditure while sitting or at rest, or 1 kcal per hour. These values were based on activities from the American Time Use Survey and listed by the National Cancer Institute. Each respondent's total outdoor physical activity was calculated by multiplying the frequency of their participation in activities with the activities' corresponding MET value annually. Table 4 indicates the average physical activity in each given exercise while Table 5 notes the MET value associated with each form

METphyaweek (Table 6)				
	Percentiles	**Smallest**		
1%	0	0		
5%	0	0		
10%	0	0	**Obs**	3,229
25%	0	0	**Sum of Wgt.**	3,229
50%	3.653846		**Mean**	8.97171
		Largest	**Std. Dev.**	14.02669
75%	12.69385	127.2712		
90%	26.67308	128.5923	**Variance**	196.7481
95%	26.67308	184.1215	**Skewness**	4.827344
99%	56.40154	278.635	**Kurtosis**	60.95234

of exercise.

Thus, the dependent variable measuring physical activity is represented by:

$$METphyaweek = (walkFreq * 3.8) + (bikeFreq * 8.0) + (sportFreq * 5.1) + \\ (swimFrq * 5.22) + (swimFrq * 5.22) \quad (6)$$

Detailed summary statistics on the composite physical activity variable, METphyaweek, are pictured in Table 6. The table finds that half of the sample engages in 3 MET of physical activity total. And a majority of that 50% of the sample reports to engage in no physical activity – even walking, with 0 MET per week. This is highly unlikely, and thus this statistic on the METphyaweek variable will be taken into account when conducting the final analysis between greenspace and physical activity.

If the hypothesis is that increased leisure physical activity and walkability is associated with an accessibility to green space, then an individuals' immediate surroundings – small geographic information is very important. Past studies that have been used to measure physical activity include National Health Interview Survey (NHIS), Behavioral Risk Factor Surveillance Survey (BRFSS), and National Health and Nutrition Examination Survey (NHNES). All three surveys examine physical activity at the national level with the lowest geographical areas at the MSA and county level, which have populations of at least 50,000. MSAs typically account for a large geographic area and are thus not good indicators of an individual's neighborhood or environmental characteristics. Thus, challenges related to the smaller sample size and non-response rate from BESTS will be adjusted in order to best account for the characteristics of the built environment.

4.3 Social Cohesion

Studies across urban planning journals note that social cohesion, or social support networks within a neighborhood, largely impact a household's behavior. Recent studies in Baltimore find a strong relationships between tree canopy and social capital (Troy et al., 2016). Thus, social cohesion may be an important indicator. The specific measurement used for social cohesion within a given geographic boundary (zip code or block group) is taken from the following dimensions: helpfulness, relationships and ties, trust, community problem solving abilities between neighbors as well as the presence of neighborhood associations, municipal services, and religious groups. These dimensions are all measured on a scale from 1 to 5 (strongly disagree, disagree, neutral, agree, strongly agree). The final dimension included asks for a number count of neighbors known by first name on a scale from 1 to 5 (none, a few, half, most, all). Assuming that these variables are measures of social capital that are free of error, the principle components analysis (PCA) is applied. In this instance, PCA is a statistical data reduction method, to minimize the number of variables in the analysis, creating a linear combination that captures the most amount of variance. In this correlation matrix, the variables have standardized unit variance with total variance of 9. Thus, the first variable, willingness to help neighbors, compromises a majority of the variation, about 3.74, which explains 46% of the total variance. More than 85% of the variance is explaining using the first five variables. The first component is used as a measurement for social cohesion.

Below are the summary statistics for the social cohesion variables. The definitions of the variables asks whether people in the neighborhood are: willing to help one another, close

Summary Statistics – Social Cohesion (Table 7)				
Variables	**Mean**	**Std Dev**	**Min**	**Max**
willHelp	3.67	1.26	1	5
closeKnit	3.42	1.34	1	5
Trust	3.52	1.31	1	5
solComProb	3.38	1.36	1	5
actNeighAs	3.67	1.53	1	5
muniServ	3.7	1.22	1	5
relGroups	3.59	1.32	1	5
knownNeighs	2.79	.98	1	5
Soccoh	0	1.93	-3.93	3

Summary Statistics – Covariates (Table 8)				
Variables	Mean	Std Dev	Min	Max
Black	.35	.47	0	1
White	.57	.48	0	1
Male	.35	.47	0	1
Education	3.125	1.24	1	5
Age	3.04	1.45	1	5
Income	3.34	2.4	1	7

knit, can be trusted, able to work together to solve community problems. The rest of the variables represent the presence of: active neighborhood associations, municipal government services, religious groups. And the final question asks participants how many neighbors they know by name.

4.4 Covariates

Variables like race, education, age, sex, income, household size, and employment status are all also control variables.

Race, sex, education, age, and income were all transformed from categorical variables into dummy variables. The majority of the samples identified as either black or white. Almost half identified as black and half identified as white. Because the sample size of the other races – Hispanic, Asian, Native American, or other – were so small, there were excluded from the analysis to prevent collinearity. Education is a categorical variable that represents the highest grade of school that an individual has the opportunity to complete. Stratified

Summary Statistics – Appended Covariates (Table 9)				
Variables	Mean	Std Dev	Min	Max
Percent White	40.6	34.5	0	98.99
Percent Black	54.5	37	0	100
Percent Other	59.37	34.5	1.01	100
Percent High School Degree	3.125	1.2	1	5
Percent Bachelor's Degree	3.04	1.45	1	5
Median Income	11	2.4	1	7
Number of Recreation Center	2.11	1.95	0	7

from 1 to 5, each level represents less than a high school degree, high school graduate, some college, college graduate, to a post-graduate degree. There were approximately 50 individuals that responded NA or refused to answer the question and these values were dropped. This variable indicates that approximately one quarter of the sample had a college degree and 15% more had a postgraduate degree. About 20% of the sample had some college experience and a little more than a third of the sample were high school graduates or less. The education categorical variable was transformed into 5 dummy variables representing each level of educational attainment.

Similarly, age is also a categorical variable, that is stratified from 1 to 5 defining under age 35, age 35 to 44, age 45 to 54, or age 65 and older. Clearly, this survey focuses on the effects of middle-age Americans, with the age group 45 to 54 being the largest; however, the large standard deviation shows that there is quite a bit of variation from age 30s to 60s. Around 20% of the sample is represented in each age bracket. Each age bracket was also transformed into five dummy variables. Finally, income is also a categorical variable stratified from 1 to 7 represented household income under $15K, from $15K to $25K, $25K to $35K, $35K to $50K, $50K to $75K, $75K to $100K, and over $100K, where each income stratification was transformed into a dummy variable.

Additional covariates in the model not only include information provided from BESTS, but also includes appended data on neighborhood characteristics from the Census. This includes data on percent of each race, gender, educational attainment, and median income at each geographic level. To control for accessibility, and account for individuals that prefer to utilize private fitness centers, data from Baltimore City Open Data on the number of recreation centers within each geographic zip code is also included. Physical activity in a public, open area is also a function of environmental characteristics like population density, weather, and crime. Because this analysis is conducted for Baltimore, weather data is omitted.

5 Empirical Model

In order to determine the effect of greenspace on physical activity and social cohesion, I applied the following statistical specification:

$$Y_{tij} = \beta_{tij} + \gamma_{tij} + \alpha_j + \epsilon_{tij} \tag{7}$$

i=1,2... N_t, each individual at a given period of time

j = 1,2 ... M, geographic boundary of each individual (zip code or census tract)

t = 1999, 2001... 2011, year

with X representing greenspace, Y representing the explanatory variables of physical activity and social cohesion, Z representing covariates, the observed neighborhood characteristics, and α_j representing unobserved, time-invariant characteristics at each geographic area, and ϵ capturing unobserved individual level characteristics that influence the outcome. In this specification, it is clear that i represents each individual that is sampled, j represents the zip code or census tract they reside in, and t represents the year; however, given that the dataset is a repeated time-series cross section, not the same individuals are sampled at different points in time. A different group of individuals is sampled each year at each geographic boundary. Additionally, this fixed effects specification is time-invariant. To control for time-variant confounders, a year dummy variable is included to monitor changes within groups across time.

In order to test which estimation method, fixed effects or random effects, should be applied, I use the Robust Hausman Test. The statistical specification assumes that there are unobserved characteristics, α_j, that are correlated with the explanatory variables. For example, characteristics like neighborhood crime rates are not included my dataset, but could be associated with greenspace. Because this violates the assumptions necessary for ordinary least squared estimates, this endogeneity can be removed through fixed effects.

Fixed effects controls for the average difference across a geographic area in both observable and unobservable variables. An example of these observable variables can be demographics, while unobservable factors can include neighborhood characteristics like walkability of a neighborhood or accessibility to grocery stores. One key assumption of fixed effects model is that unobservable factors that affect greenspace, physical activity, and social cohesion are time-invariant. The average effect of each variable is subtracted from that individual, at a certain time, within a certain geographic area, ultimately removing any unobserved factors from the final estimation on the dependent variable. A deeper analysis discussing the removal of biases can be found in Section 7. That, combined with the assumptions that there is no covariance between explanatory variables and the error term nor is there multicollinearity, is enough to assume that the fixed effects estimator is consistent.

Theoretically, fixed effects is applied when there is a single source of random variability in the data, such as a random sample. Random effects allows for random variation in other variables. A potential source of random variation in this analysis could potentially be the greenspace variable; however, as noted in the Baltimore Sustainability Plan of 2019, the City of Baltimore intentionally targets increases in greenspace in certain neighborhoods across the city, which means the greenspace variable is not randomly assigned. Within the city's Green Infrastructure initiative, the city plans to "increase green infrastructure throughout the city, targeting neighborhoods with limited access to large parks and green spaces and high disparities in health outcomes". Similarly, within the Workforce Development initiative, the city plans to "integrate and streamline the delivery of green workforce services to increase employment... and help close the equity-gap for Baltimore's low-income, African-American, and minority residents". This suggests that the city of Baltimore is already intentionally improving greenspace and green infrastructure projects in specific neighborhoods, which indicates that there is a definite relationship between unobserved factors and the effects on physical activity rates and social cohesion. Thus, this suggests that fixed effected should be applied. The results of the Hausman Test find a p-value of 0 and a high chi-squared value, 28.5, confirming the application of fixed effects.

While locational fixed effects is able to account for the estimation bias that comes from the unobserved neighborhood characteristics, nonrandom selection into neighborhoods, and greenspace bias, there are still improvements that can be made to the model. In this dataset, each individual is sampled once; however, it would be most robust to compare a single individual's changes in physical activity rates and social cohesion over a given period of time instead of comparing individuals that are similar. This geographic fixed effects is unable to account for unobservable individual characteristics that could be influencing an individuals' response to changes in greenspace. For example, two individuals with similar socioeconomic status may respond differently to changes in greenspace depending on their current state of mental health. Someone that is already depressed may perceive stronger benefits to social cohesion than someone with better base state of mental health. This limitation of location fixed effects is due to the lack of available data from BESTS.

6 Results

In the following tables, I present an analysis of the results of greenspace on physical and mental health. After scaling the greenspace variable, Tables 10, 11, and 12 can be interpreted as a single unit increase, or 1% increase in greenspace. The following results indicate a statistically significant relationship between greenspace and social cohesion, but not physical activity. Tables 10, 11, and 12 do not contain results that include all covariates, but full result tables can be found in the appendix.

The results in Table 10, indicate a small, positive association between greenspace and physical activity, but not a statistically signification relationship. The first OLS regression shows that a 1% increase in greenspace in a given zip code increases MET by 0.01, or 1% of 1 MET, which is negligible. The impact does not change in the second through fourth OLS and fixed effects results. The second OLS regression that includes all the covariates does not change the impact of greenspace on MET. When fixed effects are introduced in column 3 and 4, a 1% increase in greenspace increases MET values marginally more, to 0.1 MET, or 10% of a single MET at rest. Recalling summary statistics run on the physical activity variable suggests some form of sampling error during the sampling of the survey, since it is unlikely that half of the inidividuals perform no type of exercise per week. Because of this error in the variable itself, there is no definitive result between greenspace and physical activity.

The relationship between greenspace and social cohesion found in Table 11 at the zip code level, both OLS and fixed effects, show a statistically significant results. The initial regression found in column 1, just analyzing the association between greenspace and social cohesion, already finds a statistically significant relationship at a 1% confidence level. The OLS regressions find that a 1% increase in greenspace increases social cohesion by 0.01. Including control variables both from the BESTS dataset and appended data in the second column accounts for differences across race, gender, and age. In the second column, the correlation coefficient changes negligibly when control variables like the number of recreation centers in a zip code and demographic information are added to the regression. Finally, fixed effects is applied in columns three and four which significantly increase the effect of greenspace on social cohesion; however, when fixed effects are applied, at the zip code level, the results are only significant at the 5 and 10% confidence level and find that a 1% increase in greenspace increases social cohesion by 0.1. While the relationship between greenspace is statistically significant with fixed effects at the zip code level, it is not statistically significant at the census tract level. Despite a lack of statistical significance, the covariates respond as

The Effect of Greenspace on Physical Activity, Zip Code (Table 10)				
	(1)	(2)	(3)	(4)
Physical Activity, METphyaweek				
Greenspace	0.0123	0.0123	0.0848	0.0963
	0.014	0.0112	0.112	0.111
Year 2000	-7.411***	-7.099***	-8.009***	-8.111***
	1.269	1.274	2.212	2.206
...				
Year 2011	-14.90***	-14.14***	-14.90***	-15.24***
	1.177	1.461	1.989	2.379
Black		-0.159	0.433	0.328
		0.444	0.458	0.489
Male		0.0237	0.0341	0.0127
		0.71	0.709	0.721
Age (35 to 44)		-0.564	-0.582	-0.572
		0.831	0.812	0.83
...				
Age (over 65)		-4.460***	-4.347***	-4.456***
		0.839	0.843	0.871
Edu (hs grad)		1.554	1.428	1.52
		1.057	1.059	1.047
...				
Edu (post grad)		0.849	0.886	1.017
		0.823	0.82	0.831
Income (under 15K)		-0.923	-0.835	-1.04
		0.8	0.823	0.836
...				
Income (50 to 75K)		1.216	1.205	1.033
		1.061	1.035	0.948
Income (75 to 100K)		0.577	0.509	
		0.691	0.665	
Census Fixed Effects	N	N	Y	Y
N	3229	3157	3210	3157
Adj. R-sq	0.099	0.109	0.102	0.101

Standard errors in parentheses, * p<0.10, **p<0.05, ***p<0.01

The Effect of Greenspace on Social Cohesion, Zip Code (Table 11)					
	(1)	(2)	(3)	(4)	(5)
		Social Cohesion			
Greenspace	0.0189***	0.0159***	0.0169***	0.130**	0.132*
	0.00417	0.00318	0.00297	0.0632	0.0667
Black		-0.384***	-0.282*	-0.216	-0.22
		0.136	0.138	0.157	0.158
Male		-0.258**	-0.242**	-0.235**	-0.229**
		0.104	0.106	0.0953	0.099
...					
Age (45 to 54)		0.477**	0.440**	0.468**	0.425**
		0.183	0.181	0.177	0.177
Age (55 to 64)		0.746***	0.736***	0.764***	0.735***
		0.199	0.204	0.197	0.202
Age (over 65)		1.022***	1.037***	1.024***	1.007***
		0.191	0.198	0.188	0.195
Edu (hs grad)		0.432**	0.433*	0.414*	0.412*
		0.208	0.215	0.217	0.21
...					
Edu (post grad)		0.412*	0.382*	0.279	0.247
		0.223	0.229	0.237	0.235
Income (under 15K)		-0.404	-0.379	-0.263	-0.26
		0.267	0.27	0.274	0.274
...					
Income (75 to 100K)		-0.0418	-0.0801	-0.0242	-0.0728
		0.138	0.132	0.135	0.131
Year 2006		-0.0993	-0.12	-0.0863	-0.12
		0.11	0.109	0.1	0.101
Zipcode Fixed Effects	N	N	N	Y	Y
N	1425	1425	1385	1425	1385
Adj. R-sq	0.033	0.1	0.107	0.056	0.057

Standard errors in parentheses, * p<0.10, **p<0.05, ***p<0.01

The Effect of Greenspace on Social Cohesion, Census Tract (Table 12)					
	(1)	(2)	(3)	(4)	(5)
	Social Cohesion				
Greenspace	0.0169***	0.0156***	0.0166***	-0.00353	-0.0161
	0.00348	0.00325	0.00337	0.013	0.0151
Black		-0.452***	-0.308*	-0.0401	-0.0406
		0.155	0.159	0.217	0.219
Male		-0.347***	-0.293**	-0.328**	-0.279*
		0.128	0.133	0.143	0.15
...					
Age (55 to 64)		0.811***	0.804***	0.833***	0.785***
		0.202	0.211	0.231	0.243
Age (over 65)		1.086***	1.095***	0.970***	0.966***
		0.207	0.213	0.239	0.24
Edu (hs grad)		0.392	0.34	0.604**	0.590*
		0.25	0.248	0.281	0.284
Edu (some college)		0.169	0.103	0.216	0.22
		0.25	0.253	0.3	0.304
Edu (college grad)		0.399	0.355	0.311	0.304
		0.263	0.265	0.294	0.297
Edu (post grad)		0.457*	0.44	0.216	0.165
		0.277	0.278	0.318	0.322
Income (under 15K)		-0.433*	-0.351	-0.0899	-0.114
		0.254	0.258	0.321	0.328
...					
Income (50 to 75K)		-0.439*	-0.375**	-0.259	-0.204
		0.17	0.173	0.175	0.178
Income (75 to 100K)		-0.0245	-0.0531	-0.0709	-0.0569
		0.185	0.187	0.212	0.214
Zipcode Fixed Effects	N	N	N	Y	Y
N	965	965	939	965	939
Adj. R-sq	0.03	0.11	0.116	0.052	0.048

Standard errors in parentheses, * p<0.10, **p<0.05, ***p<0.01

expected in Table 12. For example, being male decreases levels of social cohesion. Similarly, higher education attainment and older age are also statistically significant covariates that increase the amount of social cohesion for an individual. To ensure a best, linear, unbiased estimator, OLS assumptions must be fulfilled. One of the assumptions under OLS is homoskedasticity, or that the variance of the error terms is constant. Given heteroskedastic error terms, robust standard errors are applied to the model and clustered at each geographic level.

7 Discussions

The results in Tables 9, 10, and 11 are best interpreted as the effect of a 1% increase in greenspace within a given geographic area (zip code or census tract) on physical activity and social cohesion. This interpretation is most helpful considering that organizations in Baltimore, like Tree Baltimore, are measuring variables similar to greenspace by assessing percentage increases. For example, Tree Baltimore partnered with the city to present the goal of increasing tree canopy cover, a finer measure and of greenspace, from approximately 30% to 40% by 2030. Analyzing the results from the tables will provide a better framework for understanding how a 10% increase in greenspace could potentially improve physical activity rates and social cohesion.

7.1 Physical Activity

Given the findings from Table 10 and assessment of the summary statistics, it is clear that this dataset, BESTS, is not sufficient to answer the research question and determine a causal relationship between greenspace and physical activity. The correlation coefficient has a strong weak, positive association with greenspace. Detailed summary statistics of the physical activity variable in Table 6 suggests an error with the physical activity variable itself, indicating that almost half of individuals claim to participate in 0 MET of physical activity per week.

Further, the median only participates in 3 MET per week. At the opposite end of the spectrum, the largest value of 56 MET per week is equivalent to 7 hours of biking per week. Further, analysis of the control variables also provides conflicting results that contradicts traditional theory. Given this large variability in the dataset, it is likely that there was an error in the survey sampling process. Therefore, the results of the effect of greenspace on physical activity is inconclusive.

7.2 Social Cohesion

The results of greenspace on social cohesion are more promising. The results are conducted at two different geographic boundary levels – zip code and census tract. The first column of the census tract analysis in Table 11 presents an initial OLS regression and finds a strong association between greenspace and social cohesion. Thus, interpretation of column 1 finds that a 1% increase in greenspace, is associated with a 0.019 increase in social cohesion. PCA centers the social cohesion variable around 0, however, the index for social cohesion used in the BESTS survey ranges from 1 to 5, with 1 being residents that disagree that their neighborhood has high social cohesion and 5 meaning that there are high levels of social cohesion. This suggests that a 10% increase in greenspace, proxy tree canopy cover, is achieved by Tree Baltimore, would mean an associated 0.19 increase in the social cohesion index. Column 2 of Table 11 includes covariates on individual household demographic characteristics from BESTS and finds the addition of those control variables affects the association between greenspace and social cohesion marginally, from 0.018 to 0.015, slightly dampening the impact as it controls for demographic bias. The third regression in column 3 includes covariates from appended data sources as well as BESTS and finds results comparable to the first two columns, with an increase in social cohesion of 0.017. In all three regressions, greenspace is statistically significant at the 1% level. The covariates also behave according to my hypothesis on greenspace. My hypothesis suggests that larger areas of greenspace are associated with wealthier, Whiter communities with lower crime rates, which are better environments for strong social cohesion and tighter-knit social groups. Black or minority households that are more likely to be low-income, are likely to live in areas with more crime, and thus their perception of safety and of their living environment disincentives them from utilizing their public space. Analyzing covariates finds this hypothesis holds. In columns 2 and 3, being Black decreases social cohesion by 0.038 and 0.028, respectively. The results also find that older age and higher levels of education increase social cohesion. Being 65 years of age or older, compared to being 35 years old or younger, increases social cohesion by nearly 1 unit, while high levels of educational attainment like college or post-college degrees, compared to no high school degree, increases social cohesion by nearly 0.5 units. Older age could potentially influence social cohesion and its usage in areas with more greenspace given that seniors, with more leisure time and are more influenced by accessibility to greenspace, are influencing the results. Education can play a factor, as increased education could represent decreased crime levels and perceptions of safety within one's own community. Income also behaves as expected. Compared to households earning over 100K or more, those that are low-income and earn less than 15K or 15-25K per year show nearly a 0.4 decrease in social cohesion levels. Income is also one of the only statistically significant covariates of appended data that also indicate that an increase in income is associated with a slight increase in social cohesion.

Fixed effects were applied with the original covariates from the BESTS dataset in column 4 and in column 5 with covariates from both BESTS and the appended data. Both fixed effects columns indicate large impact of greenspace on social cohesion that are both significant at a 10% confidence level. Column 4 states that a 1% increase in greenspace increases social cohesion by 0.13 and column 5 finds a 0.132 increase in social cohesion. This relationship is large – a potential 10% increase in greenspace within a zip code could increase a neighborhood social cohesion raking by approximately 1 point, from agree to strongly agree. Analyzing the covariates finds that the same relationships as previously mentioned in the OLS regression holds. Being black decreases social cohesion, which older age and higher levels of education and income increase social cohesion. While the same relationships hold, the strength of the associations are dampened. Geographic fixed effects control for unobserved characteristics within each zip code, so that holding all those features constant, can determine a causal relationship between greenspace and social cohesion.

The results from at the census tract level in Table 12, however, do not support statistically significant results as it does at the zip code level in Table 11. While a positive association still exists in the OLS regressions in columns 1-3, with a 1% increase in greenspace associated with approximately a 0.015 to 0.016 unit increase in social cohesion, similar to the OLS regression at the zip code level, the relationship is not statistically significant in columns 4 and 5 when fixed effects are applied. This could suggest that social cohesion levels at a more granular level are not directly caused by greenspace. However, another reason for this discrepancy could also be the decrease in sample size from the zip code to census tract level. Some sample in BESTS contain information on both zip code and census tracts while others only contain zip code geographies. This smaller sample size means that there are fewer degrees of freedom, which means leads to a larger test statistic and p-value. It is unclear is this data is missing at random, thus if this reduced sample size is a clear measure of the impact of greenspace on social cohesion. At this granular level, the results are inconclusive. Future studies, however, may seek a dataset with consistent geographies for its samples at increasingly more granular levels, to conclusively determine at what community size, greenspace impacts social cohesion.

The r-squared values of the results are very small, suggesting that only 5% of the variables listed capture the relationship between greenspace and social cohesion. Because the topic in question is attempting to capture human behavior, r-squared values in psychology research tends to be lower than 50% because human behavior is often more difficult to predict than physical processes. However, this could still suggest that the model provided is not a good fit to understand the effects of physical activity on social cohesion. Limitations of this model include the small sample size and delineation of geographic boundaries. These results that suggest a causal relationship between greenspace and social cohesion can be limited by the sample size of the dataset. Although the dataset is significantly smaller than those found in health economics literature, it is also larger than most datasets from urban planning journals that target this same question, which only contain 400-600 samples. Urban planning literature also raises concerns about studying trends within nationally-drawn geographic boundaries. Zip code and census tracts are arbitrary boundaries drawn by the state that facilitate data collection. The City of Baltimore has a set of geographic boundaries called Neighborhood Statistical Areas that are drawn based on the characteristics designated by natives within the city, in which each community has similar culture, values, and background that bind them together. Future studies looking to control better geographic boundaries would better represent households in those areas, by applying Neighborhood Statistical Areas in Baltimore, MD or equivalent geographic boundaries in other cities.

7.3 Estimation Bias

Traditional Ordinary Least Squared (OLS) regressions are unable to find a causal link between greenspace and health because it fails to account for sources of endogeneity, like residential self-selection, reverse causality, and non-random greenspace. These are three forms of estimation bias that can influence the effect of greenspace on health. For example, individuals who prefer to exercise could intentionally self select into neighborhoods that already have increased greenspace, or an environment that facilitates exercise, causing a bias that overestimates the effect of greenspace on physical activity. Or, reverse causality could also interfere. For example, strong social cohesion levels could encourage communities to band together to improve their own neighborhood's greenspace. Thus, the effect of social cohesion on greenspace is another example of bias that needs to be accounted for that also overestimates the benefits of greenspace. As mentioned previously, the most pertinent form of estimation bias comes from intentionally targeting increases in greenspace in certain neighborhoods, especially given that the Baltimore Sustainability Plan specifically states that the City of Baltimore intentionally targets greenspace development in low-income, African-

African, and minority residential neighborhoods. Non-random greenspace, overestimates the benefits to health, further introducing bias that traditional OLS cannot capture, making it an endogenous factor that makes it difficult to determine the causal link between greenspace and health. Applying a fixed-effects analysis at the zip code and census tract level accounts for this type of endogeneity. Fixed-effects is an estimation technique that controls time-invariant, unobserved characteristics at a given geographic area, like community culture, an individuals' perception of safety, retail environment, or features of the natural environment. After holding these unobserved variables constant and removing the overestimation bias, the results still find a strong, statistically significant relationship between greenspace and social cohesion, revealing causal link between the two.

In conclusion, the results indicate a weak positive association between greenspace and physical activity, which are ultimately inconclusive; however, the results also present a causal relationship between greenspace and social cohesion. Applying fixed effects finds that increasing greenspace by 10% can have a significant impact on social cohesion levels within a given zip code.

8 Conclusion

Using a repeated cross-sectional dataset, geographic fixed effects are applied to determine the effects of greenspace on physical activity and social cohesion, in Baltimore, MD. The results are inconclusive for the impact of greenspace on physical activity but indicate a strong, statistically significant relationship between greenspace and social cohesion. The causal relationship between greenspace and social cohesion can have policy implications that will shape future city planning, providing evidence that an increase in natural forms of the built environment can play a role in improving a neighborhood's levels of trust, how close-knit it is, and general perceptions of safety.

As urbanization trends continue, the establishment of this relationship suggests that greenspace is a critical urban design tool that can be used to improve citizen health and well-being. Singapore, internationally known as the ”garden city”, utilizes greenspace not only in the form of parks and recreation areas, but through integrated greenspace. As cities become more dense, integrating elements of the natural environment into indoor architecture, streetscapes, and green roofs become a more viable solution to allow individuals to experience the benefits of greening in high-density environments. Researchers at National University Singapore and government officials note in Singapore's Sustainability Blueprint that greenery is not only a way to achieve environmental goals that improve air quality, decrease urban heat, and manage stormwater, but is also crucial to achieving optimal individual Outdoor Thermal Comfort (OTC). Nations on the cutting edge of sustainable infrastructure find OTC as the the final frontier of well-being, that not only improves health, but comfortable environments based on temperature, humidity, shade, and vegetation.

Finally, future studies that seek to further study this relationship between greenspace and social cohesion (confirm the causal relationship found in this study), need a more comprehensive dataset that follows individuals across time and includes baseline measures of physical and mental health. Additional research needs to be conducted for cities that seek to implement greenspace by suggesting the optimal targets for greenspace. A study by Wolch et al., (2014) notes that having too much greenspace can lead to paradoxical effects like gentrification and increased property prices, which drive vulnerable individuals out of areas with better quality outdoor environments. Thus, pursuing research that finds "just enough greenspace" for each geographic area will aid policy in ensuring that target demographics, like low-income and minority residents, are able to experience improved health benefits without being priced out of a certain neighborhood.

References

[0] Astell-Burt, T., Mitchell, R., & Hartig, T. (2013). *T*he association between green space and mental health varies across the lifecourse. A longitudinal study, Journal of Epidemiol-ogy & Community Health.

[1] Barnett, D., Barnett, A., Nathan, A., Van Cauwenberg, J., & Cerin, E. (2017, August 7). *B*uilt environmental correlates of older adults' total physical activity and walking: A systematic review and meta-analysis. - PubMed—NCBI. https://www.ncbi.nlm.nih.gov/pubmed/28784183

[2] Beyer, K. M. M., Kaltenbach, A., Szabo, A., Bogar, S., Nieto, F. J., & Malecki, K. M. (2014). *E*xposure to Neighborhood Green Space and Mental Health: Evidence from the Survey of the Health of Wisconsin. International Journal of Environmental Research and Public Health, 11(3), 3453–3472. https://doi.org/10.3390/ijerph110303453

[3] Boone-Heinonen, J., Diez Roux, A. V., Kiefe, C. I., Lewis, C. E., Guilkey, D. K., & Gordon-Larsen, P. (2011). *N*eighborhood socioeconomic status predictors of physical activity through young to middle adulthood: The CARDIA study. Social Science Medicine (1982), 72(5), 641–649. https://doi.org/10.1016/j.socscimed.2010.12.013

[4] Clark, W., & Ledwith, V. (2007). *H*ow much does income matter in neighborhood choice? Population Research and Policy Review, 26(2), 145–161.

[5] de Vries, S., van Dillen, S. M. E., Groenewegen, P. P., & Spreeuwenberg, P. (2013). *S*treetscape greenery and health: Stress, social cohesion and physical activity as mediators. Social Science Medicine, 94, 26–33. https://doi.org/10.1016/j.socscimed.2013.06.030

[6] Do, D. P., & Finch, B. K. (2008). *T*he Link between Neighborhood Poverty and Health: Context or Composition? American Journal of Epidemiology, 168(6), 611–619. https://doi.org/10.1093/aje/kwn182

[7] Echeverría, S., Diez-Roux, A. V., Shea, S., Borrell, L. N., Jackson, S. (2008). Associations of neighborhood problems and neighborhood social cohesion with mental health and health behaviors: The Multi-Ethnic Study of Atherosclerosis. Health Place, 14(4), 853–865. https://doi.org/10.1016/j.healthplace.2008.01.004

[8] Eid, J., Overman, H. G., Puga, D., & Turner, M. A. (2008). Fat city: Questioning the relationship between urban sprawl and obesity. Journal of Urban Economics, 63(2), 385–404. https://doi.org/10.1016/j.jue.2007.12.002

[9] Ghimire, R., Ferreira, S., Green, G. T., Poudyal, N. C., Cordell, H. K., & Thapa, J. R. (2017). Green Space and Adult Obesity in the United States. Ecological Economics, 136, 201–212.

[10] Gordon-Larsen, P., Nelson, M. C., Page, P., & Popkin, B. M. (2006). Inequality in the built environment underlies key health disparities in physical activity and obesity. Pediatrics, 117(2), 417–424. https://doi.org/10.1542/peds.2005-0058

[11] Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. Journal of Political Economy, 80(2), 223–255. https://doi.org/10.1086/259880

[12] Hong, A., Sallis, J. F., King, A. C., Conway, T. L., Saelens, B., Cain, K. L., Fox, E. H., & Frank, L. D. (2018). Linking green space to neighborhood social capital in older adults: The role of perceived safety. Social Science Medicine, 207, 38–45. https://doi.org/10.1016/j.socscimed.2018.04.051

[13] Lachowycz, K., & Jones, A. (2011). Greenspace and obesity: A systematic review of the evidence. Obesity Reviews: An Official Journal of the International Association for the Study of Obesity, 12, e183-9. https://doi.org/10.1111/j.1467-789X.2010.00827.x

[14] Lachowycz, K., & Jones, A. P. (2013). Towards a better understanding of the relationship between greenspace and health: Development of a theoretical framework. Landscape and Urban Planning, 118, 62–69. https://doi.org/10.1016/j.landurbplan.2012.10.012

[15] Lee, A. C. K., & Maheswaran, R. (2011). The health benefits of urban green spaces: A review of the evidence. Journal of Public Health (Oxford, England), 33(2), 212–222. https://doi.org/10.1093/pubmed/fdq068

[16] Leibowitz, A. A. (2004). The demand for health and health concerns after 30 years. Journal of Health Economics, 23(4), 663–671. https://doi.org/10.1016/j.jhealeco.2004.04.005

[17] Maas, J., van Dillen, S. M. E., Verheij, R. A., & Groenewegen, P. P. (2009). Social contacts as a possible mechanism behind the relation between green space and health. Health Place, 15(2), 586–595. https://doi.org/10.1016/j.healthplace.2008.09.006

[18] Marmot, M., Wilkinson, R. (2005). Social Determinants of Health. OUP Oxford.

[19] Meyer, K. A., Guilkey, D. K., Tien, H.-C., Kiefe, C. I., Popkin, B. M., & Gordon-Larsen, P. (2016). Instrumental-Variables Simultaneous Equations Model of Physical Activity and Body Mass Index: The Coronary Artery Risk Development in Young Adults (CARDIA) Study. American Journal of Epidemiology, 184(6), 465–476. https://doi.org/10.1093/aje/kww010

[20] Morales, L. F., Gordon-Larsen, P., & Guilkey, D. (2016). Obesity and health-related decisions: An empirical model of the determinants of weight status across the transition from adolescence to young adulthood. Economics and Human Biology, 23, 46–62. https://doi.org/10.1016/j.ehb.2016.06.003

[21] Mytton, O. T., Townsend, N., Rutter, H., Foster, C. (2012). Green space and physical activity: An observational study using Health Survey for England data. Health & Place, 18(5), 1034–1041. https://doi.org/10.1016/j.healthplace.2012.06.003

[22] Nutsford, D., Pearson, A. L., Kingham, S. (2013). An ecological study investigating the association between access to urban green space and mental health. Public Health, 127(11), 1005–1011. https://doi.org/10.1016/j.puhe.2013.08.016

[23] Saffer, H., Dave, D., & Grossman, M. (2011, September). *R*acial, Ethnic, and Gender Differences in Physical Activity — Journal of Human Capital: Vol 7, No 4. https://www.journals.uchicago.edu/doi/abs/10.1086/671200

[24] Sallis, J. F., Hovell, M., Hofstette, C., Elder, J., Hackley, M., Caspersen, C., & Powell, K. (1990). *D*istance between homes and exercise facilities related to frequency of exercise among San Diego residents. - PubMed—NCBI. https://www.ncbi.nlm.nih.gov/pubmed/2108465

[25] Schüle, S. A., Hilz, L. K., Dreger, S., & Bolte, G. (2019). *S*ocial Inequalities in Environmental Resources of Green and Blue Spaces: A Review of Evidence in the WHO European Region. International Journal of Environmental Research and Public Health, 16(7). https://doi.org/10.3390/ijerph16071216

[26] Troy, A., Nunery, A., & Grove, J. M. (2016). *T*he relationship between residential yard management and neighborhood crime: An analysis from Baltimore City and County. Landscape and Urban Planning, 147, 78–87. https://doi.org/10.1016/j.landurbplan.2015.11.004

[27] Wolch, J. R., Byrne, J., & Newell, J. P. (n.d.). *U*rban green space, public health, and environmental justice: The challenge of making cities 'just green enough'—ScienceDirect. Retrieved March 20, 2020, from https://www.sciencedirect.com/science/article/pii/S0169204614000310

[28] Wolfe, M. K., & Mennis, J. (2012). *D*oes vegetation encourage or suppress urban crime? Evidence from Philadelphia, PA. Landscape and Urban Planning, 108(2), 112–122. https://doi.org/10.1016/j.landurbplan.2012.08.006

Appendix

Tables 10, 11, and 12 below are the full tables that include covariates from BESTS and appended data.

The Effect of Greenspace on Physical Activity, Zip Code (Table 10)				
	(1)	(2)	(3)	(4)
	METphyaweek			
Greenspace	0.0123	0.0123	0.0848	0.0963
	0.014	0.0112	0.112	0.111
Year 2000	-7.411***	-7.099***	-8.009***	-8.111***
	1.269	1.274	2.212	2.206
Year 2003	-4.085***	-4.156***	-5.048**	-5.217**
	1.14	1.235	2.351	2.289
Year 2006	-4.501***	-4.058***	-4.812**	-5.241**
	1.125	1.353	1.945	2.279
Year 2011	-14.90***	-14.14***	-14.90***	-15.24***
	1.177	1.461	1.989	2.379
Black		-0.159	0.433	0.328
		0.444	0.458	0.489
Male		0.0237	0.0341	0.0127
		0.71	0.709	0.721
Age (35 to 44)		-0.564	-0.582	-0.572
		0.831	0.812	0.83
Age (45 to 54)		-2.334***	-2.292***	-2.327***
		0.792	0.766	0.775
Age (55 to 64)		-2.296***	-2.177***	-2.333***
		0.76	0.754	0.747
Age (over 65)		-4.460***	-4.347***	-4.456***
		0.839	0.843	0.871
Edu (hs grad)		1.554	1.428	1.52
		1.057	1.059	1.047
Edu (some college)		0.263	0.214	0.241
		0.904	0.896	0.888

Edu (college grad)		0.585	0.72	0.698
		0.802	0.835	0.801
Edu (post grad)		0.849	0.886	1.017
		0.823	0.82	0.831
Income (under 15K)		-0.923	-0.835	-1.04
		0.8	0.823	0.836
Income (15 to 25K)		-0.319	-0.281	-0.449
		0.774	0.773	0.778
Income (25 to 35K)		-0.647	-0.747	-0.839
		0.827	0.859	0.871
Income (35 to 50K)		0.243	0.334	0.19
		0.936	0.946	0.912
Income (50 to 75K)		1.216	1.205	1.033
		1.061	1.035	0.948
Income (75 to 100K)		0.577	0.509	
		0.691	0.665	
HS Degree		-0.00716		-0.00313
		0.00389		0.00814
College Degree		0.00531		0.00777
		0.00381		0.00562
Median Income		-0.0056		-0.00339
		0.00435		0.00765
Census Fixed Effects	N	N	Y	Y
N	3229	3157	3210	3157
adj. R-sq	0.099	0.109	0.102	0.101

Standard errors in parentheses, * p<0.10, **p<0.05, ***p<0.01

The Effect of Greenspace on Social Cohesion, Zip Code (Table 11)

	(1)	(2)	(3)	(4)	(5)
		Social Cohesion			
Greenspace	0.0189***	0.0159***	0.0169***	0.130**	0.132*
	0.00417	0.00318	0.00297	0.0632	0.0667
Black		-0.384***	-0.282*	-0.216	-0.22
		0.136	0.138	0.157	0.158
Male		-0.258**	-0.242**	-0.235**	-0.229**
		0.104	0.106	0.0953	0.099
Age (35 to 44)		0.332*	0.288	0.357*	0.322*
		0.178	0.187	0.178	0.177
Age (45 to 54)		0.477**	0.440**	0.468**	0.425**
		0.183	0.181	0.177	0.177
Age (55 to 64)		0.746***	0.736***	0.764***	0.735***
		0.199	0.204	0.197	0.202
Age (over 65)		1.022***	1.037***	1.024***	1.007***
		0.191	0.198	0.188	0.195
Edu (hs grad)		0.432**	0.433*	0.414*	0.412*
		0.208	0.215	0.217	0.21
Edu (some college)		0.222	0.209	0.192	0.181
		0.194	0.199	0.191	0.186
Edu (college grad)		0.489*	0.505*	0.415	0.412
		0.247	0.252	0.252	0.251
Edu (post grad)		0.412*	0.382*	0.279	0.247
		0.223	0.229	0.237	0.235
Income (under 15K)		-0.404	-0.379	-0.263	-0.26
		0.267	0.27	0.274	0.274
Income (15 to 25K)		-1.113***	-1.103***	-1.036***	-1.071***
		0.28	0.289	0.299	0.307
Income (25 to 35K)		-0.183	-0.161	-0.0809	-0.0847

		0.199	0.207	0.194	0.201
Income (35 to 50K)		-0.268*	-0.237	-0.175	-0.178
		0.153	0.154	0.162	0.161
Income (50 to 75K)		-0.343**	-0.315**	-0.326***	-0.305***
		0.106	0.104	0.109	0.108
Income (75 to 100K)		-0.0418	-0.0801	-0.0242	-0.0728
		0.138	0.132	0.135	0.131
Year 2006		-0.0993	-0.12	-0.0863	-0.12
		0.11	0.109	0.1	0.101
Number of Rec Centers			0.0396		0
			0.0418		(.)
HS Degree			-0.00027		-0.00015
			0.00115		0.00304
College Degree			0.000388		-0.00614*
			0.00147		0.00301
Median Income			0.0026		0.00182
			0.00167		0.00265
Percent Other			-0.00704		-0.0327*
			0.0133		0.0134
Percent Black			0.00227		0.027
			0.0123		0.0133
Zipcode Fixed Effects	N	N	N	Y	Y
N	1425	1425	1385	1425	1385
Adj. R-sq	0.033	0.1	0.107	0.056	0.057

Standard errors in parentheses, * p<0.10, **p<0.05, ***p<0.01

The Effect of Greenspace on Social Cohesion, Census Tract (Table 12)

	(1)	(2)	(3)	(4)	(5)
		Social Cohesion			
Greenspace	0.0169***	0.0156***	0.0166***	-0.00353	-0.0161
	0.00348	0.00325	0.00337	0.013	0.0151
Black		-0.452***	-0.308*	-0.0401	-0.0406
		0.155	0.159	0.217	0.219
Male		-0.347***	-0.293**	-0.328**	-0.279*
		0.128	0.133	0.143	0.15
Age (35 to 44)		0.385*	0.32	0.404*	0.33
		0.203	0.208	0.227	0.232
Age (45 to 54)		0.569***	0.499**	0.600***	0.565**
		0.187	0.194	0.216	0.228
Age (55 to 64)		0.811***	0.804***	0.833***	0.785***
		0.202	0.211	0.231	0.243
Age (over 65)		1.086***	1.095***	0.970***	0.966***
		0.207	0.213	0.239	0.24
Edu (hs grad)		0.392	0.34	0.604**	0.590*
		0.25	0.248	0.281	0.284
Edu (some college)		0.169	0.103	0.216	0.22
		0.25	0.253	0.3	0.304
Edu (college grad)		0.399	0.355	0.311	0.304
		0.263	0.265	0.294	0.297
Edu (post grad)		0.457*	0.44	0.216	0.165
		0.277	0.278	0.318	0.322
Income (under 15K)		-0.433*	-0.351	-0.0899	-0.114
		0.254	0.258	0.321	0.328
Income (15 to 25K)		-1.223***	-1.175***	-0.733**	-0.762**
		0.267	0.271	0.335	0.337
Income (25 to 35K)		-0.183	-0.146	-0.0407	0.00462
		0.223	0.229	0.263	0.273

Income (35 to 50K)		-0.277	-0.208	-0.269	-0.224
		0.215	0.216	0.246	0.245
Income (50 to 75K)		-0.439*	-0.375**	-0.259	-0.204
		0.17	0.173	0.175	0.178
Income (75 to 100K)		-0.0245	-0.0531	-0.0709	-0.0569
		0.185	0.187	0.212	0.214
Number of Rec Centers			0.0117		0.127
			.0377		0.093
HS Degree			-0.000751		-0.00179
			0.00156		0.00283
College Degree			-0.00185		-0.00127
			0.0015		0.00274
Median Income			0.00285		0.00203
			0.0018		0.00342
Percent Other			-0.00337		0.00892
			0.0133		0.0302
Percent Black			-0.000681		-0.00976
			0.0123		0.028
Zipcode Fixed Effects	N	N	N	Y	Y
N	965	965	939	965	939
adj. R-sq	0.03	0.11	0.116	0.052	0.048

Standard errors in parentheses, * p<0.10, **p<0.05, ***p<0.01